The HIGH SCHOOL JOURNAL

Vol. XLI December, 1957 No. 3

IN THIS ISSUE

The Use of Tests in Secondary Schools.....*C. M. Merritt* 66

Who Benefits from High School Testing
Benno G. Fricke and Jason Millman 71

Do Students Benefit from Testing?....*J. Wayne Wrightstone* 75

Testing as a Teaching Device.....*Nicholas A. Fattu* 79

Teachers Don't Know How to Test.....*John A. Green* 83

Measure—Evaluate—Then, Adapt
Vivien Ingram and Susan Kilmer 87

Can We Measure What We Teach?.....*John C. Flanagan* 93

Are Teachers Qualified to Develop, Administer and Interpret Tests?.....*Dugald S. Arbuckle* 97

Sociometric Measurements in the School....*Walter D. Smith* 103

Recent Books in Measurement.....*William P. Matthews* 108

THE CONTENTS OF THIS HIGH SCHOOL JOURNAL ARE LISTED IN "EDUCATION INDEX"

Issued Monthly, October to May, Inclusive $2.00 a Year

Published by the University of North Carolina Press for the School of Education of the University of North Carolina.

SAMUEL M. HOLTON, *Editor*

Editorial Office: Box 810, Chapel Hill, N. C.

Business Office: Box 510, Chapel Hill, N. C.

The Use of Tests in Secondary Schools

C. B. MERRITT
College of Education, University of Arizona

THE MAIN OBJECTIVE of all educational endeavor is to promote learning. Ultimately, the worth of any activity carried on in the schools has to be judged by the contribution this activity makes toward the improvement of learning. Testing is no exception to this and neither is administrative work, supervision, or any other service performed by school personnel. In a broad sense, desirable learning has taken place when behavior has been changed in socially approved directions. Testing is justified only if it contributes to our efforts to bring about such changes. To a degree, measurement procedures are now relied upon to provide descriptions of those aspects of human behavior which are of direct concern to educators. No one questions the desirablity of accurate information about student aptitudes and achievements. The most serious indictments of tests concern the accuracy and completeness of the descriptions provided, not the desirability of such information. The purpose of this paper is to review some of the functions of measurement at the high school level and also to point out certain of the misuses of tests and test results. The discussion is aimed primarily at the use of standardized tests although much of it is equally appropriate in connection with informal, teacher constructed tests.

Standardized tests suitable for administration to groups of students have been widely used for about forty years. Millions of such devices are purchased and used each year. Although the cost per student is relatively small, the total amount spent for tests runs into the hundreds of thousands of dollars yearly. How well are test results used? A definitive answer to this question is impossible. No one person can know what use is made of test data in all or most of the nation's schools. There are many legitimate uses for the results of measurement but gross abuses are not unknown. It seems obvious that some schools and some teachers make better use of test results than do others. The remainder of this discussion is devoted to factors which determine the extent to which tests are well used. The organization of the material is indicated by the headings and sub-topics.

PREREQUISITES TO THE EFFECTIVE USE OF TEST RESULTS

Perfect tests, administered under ideal conditions and interpreted wisely, would make this particular discussion quite unnecessary. Since such ideal conditions do not prevail, it is pertinent to review the prerequisites which should be satisfied before test results can be effectively used.

Cooperation with test producers. Cooperation between teachers and the makers of standardized tests is especially necessary in the construction of achievement tests. Test makers need to be advised as to the objectives and content which teachers consider basic. To be effective at all, achievement tests must reveal the extent to which students have progressed toward the attainment of important educational goals. One of the primary critiscisms of standardized achievement tests is that they overstress the ability to memorize facts and fail to measure other more complex but more important abilities. Teachers could be of real help if more frequently allowed to offer suggestions and criticisms in the preliminary stages of test contruction. The building of standardized achievement tests, especially test batteries, now involves so many technical complexities that the job is probably best left to agencies which can muster people with the necessary skills. This does not mean however that the measuring instruments developed should not measure progress toward objectives which represent what the best teachers consider to be important goals in their own areas.

The training of test users. A second prerequisite to the effective use of test results concerns the training of those who interpret such results to students and to themselves. A test score is simply the numerical description of student performance on a representative sample of questions from a given area. Since the items are but a sample and since the test only partially reveals the student's real abilities, scores must be interpreted with caution. Any teacher or counselor who works with tests should have an understanding of the few elementary satistical concepts necessary to properly interpret a score. Such concepts as mean, median, percentile, standard score, correlation, and standard error of measurement are not too difficult to master sufficiently to help one avoid misusing or even misunderstanding a score. The better test manuals now give a wealth of information about the tests which they accompany. This material can neither be understood nor used intelligently without a deliberate attempt to master the techniques involved.

The testing program. Another prerequisite to the effective use

of tests concerns the nature of a testing program. Here the continuity of the program is important. Are scores comparable from year to year so that progress may be evaluated? If so, it usually means that the same or equivalent tests are given for a period of years. New tests are introduced only when it is clear that they are superior to those in use or when it is desired to measure some new characteristic. When more than one type of test is used, care should be taken to insure that the norm groups are comparable. The modern trend here is toward batteries of tests, all standardized on the same group. Several types of tests are usually desirable in a testing program. Achievement tests supplemented with aptitude tests usually lead to a more adequate interpretation of both.

The administration of tests. Test scores are useful only when the tests have been carefully administered, accurately scored, and the results made available in understandable form. Since test scores are approximations at best, it is extremely important to obtain such scores under the best possible conditions and perform the clerical work accurately. Even this is of little help until the raw scores have been converted to meaningful derived scores. Because test scores are often useful over a period of years, some form of cumulative record must be maintained.

The question of validity. If perfect tests were available, the work of teachers would be greatly simplified. Individual differences could be clearly identified, students' strengths and weaknesses would be made obvious, and the pattern of each individual's development could be known with confidence. Unfortunately this is not the case. Consequently the most important reason for our not being able to use tests effectively is that the measuring instruments now available are only partially truthful indicators of the characteristics we would like to measure. In the face of this situation there are several courses of action possible. Some say we ought to throw out the tests. This is not a good answer because the need for even partially truthful information is too insistent. The naive worker goes ahead and acts as though test scores were perefectly valid indicators of the characteristic named in the title of the test. The perfect example of this is the teacher who would say: "I know exactly what Johnny can do because I gave him an IQ test and his was 109." Such an approach to test interpretation leads to gross oversimplification of the problems involved. A more reasonable approach is to grant that improvement in tests are necessary but to go ahead and learn what the deficiencies of particular tests are and what precautions are necessary for their proper interpretation.

Valid tests seem more possible in the area of school achievement than in the areas of aptitude or personality measurement. In constructing a test to measure achievement in some academic field, the initial step is to define what it is that is to be measured. Good progress has been made recently in defining the skills and behaviors which can reasonably be expected from students who have mastered the general objectives in an academic area. Broad goals are stated first, then each is defined in various ways to spell out the behaviors which can be expected of students who have mastered the broad goals. The final step is to develop test items which will reveal whether or not the student possesses the desired behavior. The hard job is to define objectives specifically enough to be of real help in test construction.

The use of tests could be improved if educators were more insistent that standardized tests be developed in accordance with the steps outlined in the paragraph above. Too often, standardized test-makers start by determining the information common to various textbooks used for a particular course. Since memory for specific items of content is more easily measured than complex but more important behaviors, such tests often turn out to depend too much on memory for factual information. This is due in part to the fact that teaching, for some, consists largely of the presentation of facts to be learned by students. It is also due, in the writer's opinion, to the fact that it is relatively easy to develop test items to measure memory for factual information. This behavior (the ability to recall specifics) is important to be sure but is not the only desirable outcome of learning. This particular criticism of tests is most deserved by some of the older measuring instruments. Some recent tests do successfully measure more complex behaviors.

SOME PARTICULAR USES OF TESTS

If measurement is to contribute importantly to the improvement of the educational process, the results of measurement must be used to assist us to re-define goals, modify precedures, and examine in general our total educational effort. It is one thing to give tests to determine the abilities of students. It is quite another to use this information as the basis for a re-examination of the activities carried on under the name of education. Test results can provide useful information but they cannot provide the judgments which need to be made on the basis of such information. Two ways in which test results might be better used are discussed below.

The use of tests to improve instruction. Some rather obvious

possibilities may be mentioned here. Teaching certainly will be most effective when best adapted to the abilities of the learner. Measurement procedures can reveal these abilities, at least to a degree. Knowing the extent of individual differences in aptitudes and level of achievement, teachers can modify objectives, procedures, and content to better meet the needs of each individual. Accurate description of individuals is desirable but effective use of this information to improve classroom procedures is even better. This is one way in which tests can be used to better advantage than is commonly done.

Test results may be used to determine the effectiveness of the instructional procedures used. Any concern turning out a product is interested in the quality of that product. The products of our schools are students who leave, we hope, with improved skills and abilities of various kinds. We should know a great deal more than we do about the effect instructional procedures have had upon the course of learning. Complete satisfaction with current procedures prevents the use of tests for this purpose.

The use of tests in counseling students. The use of tests for guidance purposes is generally undertaken to aid students in realistically appraising themselves so that sound decisions as to courses of action may be made. Decisions based on accurate information are certainly desirable. The difficulty here concerns the accuracy and completeness of the information obtained from tests rather than the desirability of this information. The differential prediction of success in various occupational fields and areas of school work is now partially possible. The temptation is great, however, to overdo the accuracy of such predictions. Students sometimes leave the counselor's office with the firm conviction that test scores have guaranteed their success in some occupation or college curriculum. Such predictions are best couched in terms of probabilities or expectancy tables. Flat predictions of success are just not possible.

Another pitfall in this area concerns the interpretation of any single individual's score on a test. Some test users assume that such a score is a very accurate and precise thing. This view fails to recognize that test scores are approximations only. While this is not the place for a lesson in statistics, there are techniques for evaluating the reliability of an individual score. The better test manuals report the standard error of measurement for the test which they accompany.

Who Benefits from High School Testing?

BENNO G. FRICKE
Evaluation and Examinations Division
JASON MILLMAN
Research Assistant, University of Michigan

IN ATTEMPTING to answer the question "who benefits from high school testing?" for the issue on "Measurement and Evaluation" it became readily apparent that the objective assessment of who benefits would be almost impossible. Empirical data to answer the question simply are not available. Perhaps it would be possible to assemble arguments and evidence to support the view that no one benefits very much from high school testing, or to support the view that everyone benefits a great deal from high school testing, but it is probable that all extreme answers would be wrong by quite a bit. In any event, who benefits, as well as the extent of the benefits, will depend to a great degree upon the nature of the testing program.

Perhaps a closer look at some of the people who can benefit from the testing program and the nature of the testing program by which this can come about will be worth while.

Since the student is the central element in the educative process, it is obvious that the student should benefit from the testing program. And he can benefit if proper use is made of the test reults.

Accurate assessment of a student's ability and personality traits may result in providing the student with a profitable and happy educational experience in high school. Much depends on the school's ability to recognize and handle differences among students. Since the material considered in a typical high school class is usually too difficult for the weak students and too easy for the strong students, some form of homogeneous grouping is usually beneficial. Another way in which a testing program can benefit a student is by guiding him toward an educational and vocational objective in line with his ability and personality. An opportunity for success and satisfaction in the future can be planned for, and with, each student.

But unfortunately not all students benefit from high school testing. All too often test results are filed away in a cabinet, imbedded in a folder, and are not seen by students nor used by the school staff. In this case the test publisher is the only one who benefits!

As a result, many students miss one of the best opportunities for gaining a realistic understanding of themselves. Use of a test profile to graphically portray each student's traits, abilities and achievements renders lengthy and often confusing verbal explanations unnecessary. In many instances students require only a minimum of assistance in arriving at accurate inferences and conclusions.

Not only do some students not benefit from the testing program, but some may actually be harmed from the improper use of test results. Important decisions about, and by, students are sometimes made on the basis of tests having little or no validity of any practical importance. While the research evidence on the utility of projective personality tests such as the Rorschach, Thematic Apperception Test, and others continues to be mostly negative[1,2] a few overly enthusiastic school and counseling psychologists draw unwarranted conclusions about students. The situation with respect to objectively scored personality inventories is not much better.[3] It is the opinion of the authors that the systematic use of anecdotal records, sociometric ratings, and similar non-test devices[4] by conscientious and guidance-orientated teachers are more valuable in identifying students having an undesirable or unhealthy personality.

Errors are also frequently made in the use of tests of aptitude, ability, and achievement which are usually much more valid than personality tests. Again results are often not considered as estimates of aptitudes abilities, etc., but are all too often considered as final and unalterable facts about the student. They are taken as "dictators" rather than "indicators." Some school and counseling psychologists, teachers, and principals overweight the significance of high test scores and underweight the significance of low scores. For example, in general, bright students who have performed in high school below their tested ability level (underachievers) are likely to do poorly in most colleges. It is difficult for a person giving guidance to such a bright student to avoid expressing optimistic forecasts but statistical evidence is clear on this point.

Thus students can benefit from the testing program, but there are many precautions to be heeded. From the foregoing discussion implications concerning others who may benefit from the test-

[1] Kelly, E. Lowell. Theory and Techniques of Assessment. Pages 281-310 in *Annual Review of Psychology*. Stanford: Annual Reviews, Inc., 1954.
[2] Cronbach, Lee J. Assessment of Individual Differences. Pages 173-196 in *Annual Review of Psychology*. Stanford: Annually Reviews, Inc., 1956.
[3] Furst, Edward J., & Fricke, Benno G. Development and Applications of Structured Tests of Personality. *Review of Educational Research* 1956, 26, 26-55.
[4] Froehlich, Clifford P., & Darley, John G. *Studying Students: Guidance Methods of Individual Analysis*. Chicago: Science Research Associates, 1952.

ing program suggest themselves. A brief comment on some benefits accruing to others follows.

The teacher benefits insofar as he has a better understanding of the students in his class. Test results can help him to deal more effectively with them. Achievement tests probably find their greatest use with teachers. Counselors, and other guidance workers attempting to assist students with their personal, educational, and vocational-choice problems benefit from test results not only because they provide quick and economical descriptions of students but also because they tend in general to be more reliable and valid than other means of assessing students.[5]

By comparing test results of students in his school with national norms the school administrator can objectively determine the areas of strength and weakness. But since standardized achievement tests do not take into account local objectives and problems, an evaluation of the curricula or teachers by this method alone would be unwise.

In a subtle or direct manner parents always put some kind of pressure on their children concerning school achievement and educational or vocational choice. Since many parents tend to think of their children as being "really" above average, or use them as a means by which their own wishes can be vicariously fulfilled, their expectations often may be too high or in the wrong direction. When large discrepancies between parents' expectations and tests results appear, parents should be helped to understand their children more objectively. Most parents are able to benefit from a sincere and enlightened discussion of their child's potentialities, interests, and characteristics.

Colleges benefit whenever they receive a student who "belongs." About half of the students who go to college do not graduate. Test scores have not been used optimally in the determination of who is qualified for college or who should be guided to what college to achieve the greatest benefit.[6,7] Employers benefit when sophisticated use of test scores is employed in deciding on an appropriate occupational activity for each student.[8] Carrying this to its logical

[5] Meehl, Paul E. *Clinical Versus Statistical Prediction: A Theoretical Analysis and a Review of the Evidence.* Minneapolis: University of Minnesota Press, 1954.

[6] Fricke, Benno G. Prediction, Selection, Mortality, and Quality Control. *College and University: Journal of the American Association of Collegiate Registrars and Admissions Officers* 1956, 32, 34-52.

[7] Fricke, Benno G. Tests, Students, Standards, Admissions Officers (in press) in *Long-Range Plans for Higher Education.* Washington: American Council on Education, 1958.

[8] Super, Donald E. *Appraising Vocational Fitness by Means of Psychological Tests.* New York: Harper and Brothers, 1949.

conclusion when more efficient use is made of our human resources (and there is an urgent need for this now), the nation benefits.

When a student who, because of proper use of test results, is well adjusted and challenged in his school classes, happy with his curriculum, and aware of his abilities and interests with respect to his educational and vocational future, then not only he himself benefits, but in the long run, teachers, counselors, schools administrators, college personnel, employers, and others benefit.

Do Students Benefit from Testing?

J. Wayne Wrightstone
Director, Bureau of Educational Research
Board of Education of the City of New York

THE QUESTION asked is: Do students benefit from testing? A detailed answer to this question may be found in a booklet entitled, *What Tests Can Tell Us about Children.*[1] In this article, only a briew summary of some of the benefits, both direct and indirect, can be presented regarding the use of standardized and unstandardized tests and techniques. There are various categories of tests from which students may benefit. For convenience in this discussion, the four major classifications are: (a) intelligence or academic aptitude tests; (b) special aptitude tests, for example, in music, art, mechanical, clerical, and other types of activities; (c) achievement tests of concepts, information, understandings, and principles of various school subjects or courses; and (d) assessment of personal and social adjustment.

Tests of Special Aptitude

Tests of special aptitudes help students and their parents to decide, for example, whether or not a student should have music lessons, based in part on results from the administration of a musical aptitude test. This test will measure the student's sense of pitch and time, tonal memory, ability to distinguish between melodies, and sense of rhythm. Similar special aptitude tests may be used to measure a student's capacity or talent for such activities as art, science, mechanical, clerical, and stenographic activities.

Tests of Intelligence of Academic Aptitude

The tests of intelligence, or academic aptitude, provide benefits for the student, parent, and the teacher as an aid for estimating the kinds of school courses that the individual should choose or the general classification of occupation for which the individual might prepare. It also provides a measure to help in judging whether or not the student may be achieving in school subjects in accordance with his potential academic ability. It is true that test results do not guarantee success. An extremely bright high school student may fail in college because he lacks the motivation, interest and drive to succeed. On the other hand, students of average or

[1] Published by Science Research Associates, Inc., 57 W. Grand Ave., Chicago 10, Ill.

even below-average mental ability sometimes master college work because of high motivation and hard study.

TESTS OF ACHIEVEMENT

Tests of achievement are particularly valuable because they help the student and teacher to know his status in such subjects as reading, English, mathematics, social studies, science, or foreign language. This permits the teacher to adapt the curriculum to the achievement level of the students. The achievement tests show the teacher and student the relative amount of facts, skills, and understanding that have actually been mastered. In this way, they differ from aptitude tests which provide an estimate of the potential ability of a student.

PERSONALITY TESTS

Personality measures are usually regarded as tests, but are generally referred to as questionnaires, inventories, schedules, or records. If a student is given to frequent moody spells, the teachers may wonder if he has a normal personality. This is not an easy question to answer even though experts have designed some tests to help teachers find out. When educators speak of personality, they are referring to a combination of a person's interests, attitudes, ways of behaving, and his temperament. Although there is no one test that can accurately measure all aspects of personality, teachers can find out through interest inventories, attitude scales and personality inventories, the kinds of activities that a student likes or enjoys, the kinds of moods he says he experiences in different situations, and whether he prefers to be with other people or by himself.

INFORMAL TESTS AND TECHNIQUES

Students benefit from informal evaluation techiniques in addition to formal, or standardized, tests. These are described in a book by Wrightstone, Justman, and Robbins, entitled *Evaluation in Modern Education*.[2] Among the various informal techniques that are used the following may be briefly described:

(a) Observing the student at work, which may be done by both informal observation or by means of formal observation checklists which indicate the particular kinds of behavior that are to be observed.

(b) Rating scales are similar to the observation checklist but they permit teachers to make judgments about the degree

[2] Published by American Book Company, New York City, New York.

to which a certain characteristic is present in a person. Usually the rating scales are used to judge or estimate such types of behavior as leadership, industry, sociability, and effect upon other students.

(c) Anecdotal records also provide teachers with observations that may benefit the student by permitting them to guide his behavior more wisely. The anecdotal records are notes about the pupil compiled over a long period of time. The teacher records pertinent situations, activities, experiences, and statements of the students. Experienced teachers use these records to help them evaluate such characteristics as emotional adjustment and social adjustment.

(d) Sociometric tests are also used to help the pupils. These tests do not have norms and are not standardized. In a sociometric test each individual is asked to name the person with whom he would like to associate in a common situation or activity. When the choices have been made and are diagrammed in graphic form, teachers have a picture called a sociogram. This shows which pupils are mutual friends, which are in cliques, which are isolates, and which are popular "stars."

(e) The interview is also an excellent way of finding out information for counseling a student. The interviewer generally makes some sort of outline in which he fills out certain notes on the data that he has obtained from the pupil. The interview permits a deeper analysis of the individual's difficulties, problems, strengths, and weaknesses.

How Tests Help Students

The tests, both teacher-made and standardized, generally benefit pupils by helping to pinpoint the differences among pupils. They show us the specific ways in which children differ from one another and also the degree to which they differ. This permits adaptations of instruction to individual differences among pupils.

The tests help teachers to help students by matching achievement with ability. Thus, they are able to determine, for example, whether John has a potential ability to read better than he is actually reading at the present time. One of the most practical school uses of aptitude and achievement tests is to tell us whether a student is working up to his level of ability.

Tests also help us to accept a student's limitations. Wise parents

or teachers will accept the student's limitations in academic aptitude or achievement. They will not set unrealistic goals that will cause him frustration. Tests help the child to accept himself as he potentially is and to make the most of whatever abilities he does have.

The tests also help teachers to change what can be changed. Tests can help teachers to discover difficulties that can be remedied. These defects may be physical, for example, defects of vision or hearing, or they may be emotional such as disturbed personality, immaturity, or lack of interest, or they may be instructional as deficiency in reading or mathematics skills. The tests help us to help students by reinforcing the strong points that a student may have. This is equally important with the use of tests to find out deficiencies or limitations that can be remedied.

Testing as a Teaching Device

NICHOLAS A. FATTU
Institute of Educational Research, Indiana University

EXPERIENCED TEACHERS have long recognized that testing was one of the most effective ways of insuring that a given educational objective would be mastered. Testing can help the teacher by giving him a means of effective study of the instructional process —its goals and its outcomes. Testing also can help the student by clarifying what is to be learned and by indicating the nature and extent of his progress in the desired direction.

The example below indicates how testing produced significant improvements in teaching and learning. To do so, testing went hand in hand with instructional planning, and the results of testing were constantly used to direct the planning of the course offerings.

Example:

Basic Engineering, a course for Navy enlisted men, was designed to prepare men for performance of various tasks in the engine room of a ship. Three-sevenths of the time in the course was devoted to classroom work and four-sevenths to shop work. Instructional units, requiring from one to five weeks, covered such equipment as diesel engines, refrigerators, boilers and turbines, pumps and valves, etc. Before intensive work in testing was started, it was noted that graduates seemed well-prepared in their academic work but lacked ability to do the jobs that required the technical "know-how" they had presumably learned in the shop. For instance, on a typical unit, diesel engines, it was observed that students could pass a pencil-paper test, but seemed to fail far too often on such tasks as operating the engine properly, replacing and timing an injector, etc. A comprehensive testing program was set up, as one of the tasks under NDRC Project N-106. Besides standardized pencil-paper tests of classroom performance, an intensive effort was made to develop performance tests of the shop skills. All tests were developed through the joint efforts of a test specialist and the instructors in the course.

As performance tests were developed and used in one instructional unit after another, an interesting trend was observed.[1] When

[1] These data are taken from the author's notes. Some of the data are referred to in Gulliksen, H., "Intrinsic validity," *American Psychologist* 5:511-517, 1950; Stuit (40);

the performance test was first used, the average score was about 20 to 25 points out of a possible 100. As each successive class passed through the instructional unit, the scores continued to go up, until they stabilized (in the fifth to seventh successive class) at an average level of about 80 to 90 points. Thus trainees, in the sixth class through an instructional unit typically averaged 300 per cent or higher scores than those in the first class!

Since the performance tests covered the tasks that students had to perform later on the job, this meant that they were correspondingly better prepared for their jobs.

It is interesting to note that classroom learning did not suffer during this period. In one school, on a comprehensive standardized pencil-paper test of classroom learning, scores increased from a mean of 70.3 (before improved testing) to 105.5 (after). In another school the mean scores were 61.3 (before) and 101.5 (after).

It has been indicated that tests can have a powerful influence on learning if they are designed to attain the same goals as the teaching process. It should be added that testing can also produce undesirable results when badly used. The New York State Regents examinations, for a while, had the unfortunate effect of turning secondary schools into cramming schools directed toward the probable content of the next examination as revealed by past examinations. Such unwarranted control over the teaching process defeats the purpose of the examination and also puts a hurdle in the way of education itself.

Tests are effective teaching devices only when they are used as an integral part of the teaching process. When the teaching is planned, the measurement should be planned. The testing program must be as closely related to the objectives you are trying to achieve as is the instruction. Under these conditions tests can help the teacher in several steps of the instructional program:

1. They can help clarify the objectives to aim at or the changes to seek in students.
2. They can help determine what content and learning experience can best be used to attain these objectives.
3. They can help determine an effective organization of the learning so that its effect will be cumulative.
4. They can help to appraise the effects of learning experiences —to find out where they have been effective and where they have not produced the results desired.

Porter, R. B., and Fattu, N. A., "Test development in Basic Engineering," *Proceedings of the Joint Meeting of the Bureau of Naval Personnel, Standards and Curriculum Division, and NDRC Project N-106*, February 1945, pp. 70-75.

Although only step four is commonly associated with measurement, it is not possible to measure unless the major objectives have been identified *specifically* and defined in terms of *specific* phases of pupil behavior—knowledge, information, skills, interests, work habits, etc.

Instruction is most efficient when it is directed. An instructor needs to determine the objectives he aims at and to select those that can be reasonably attained.

Although careful selection and specification of objectives is recognized as important, actually many schools and teachers carry on their instruction without having a clear outline of the goals to be attained. Many find themselves using a particular instructional procedure or book because they have used it many years and accept it as useful without any clear notion of the underlying purpose.

It is entirely possible to carry on an instructional program for years without clearly specifying objectives. It is *not* possible to construct a valid achievement test, or to use one properly, without definitely clarifying the objectives that the test is to measure.

This imperative need of testing to have objectives stated in specific terms, rather than vague generalities, can stimulate the teachers to study the objectives and carry the analysis to a degree of definiteness that could not occur if the testing problem were not present. Educational measurement can help in selecting and clarifying educational objectives by stimulating the faculty to express these objectives clearly in terms of desired behaviors. When teachers focus their attention specifically on the goals, surprising clarification can arise. This improves instruction by giving it a definite goal rather than a vague one.

Testing also provides a means for obtaining data about student performance. These data can be used for making a rational choice of objectives. Testing helps in selecting content and learning experiences. By indicating level of student attainment, test data provide a basis for noting the rate at which students attain the objectives. This suggests the levels of content, instructional materials and procedures that are likely to be effective. Analyses of test results can indicate if failure to attain objectives was due to poor material or procedures. Test results thus provide a rational basis for continued improvement of course content and learning experiences.

In organizing learning experiences, the intent is to combine them so that their cumulative effect will be the greatest possible.

Testing contributes to organization by stimulating teachers to inquire into the bases of organization. As with objectives, it is quite possible to teach without being aware of the organization in the program one is teaching. However, it is not possible to construct an effective comprehensive achievement test in a course without specifying the organization. This is a significant achievement, because one's organization of a course will not be improved, unless it is explicitly identified by the teacher himself.

Testing also can contribute to organization of instruction by providing a means of checking effectiveness of various alternatives in organization. This can separate those that are effective, and call attention to those that do not work.

Testing can thus contribute to the education of teachers in service. It stimulates the development of understanding of teachers by giving them a means of effective study of the instructional process and its outcomes.

Tyler has indicated that

> "Many of the present inadequacies of instruction are due to the common tradition in many schools of making no significant changes in the instructional program over the years, or the equally unintelligent practice in many other schools of making changes on the basis of current fads or personal whims of the staff.
>
> . . . Educational measurement can have a profound influence in the improvement of instruction; but to do so, it must be viewed as an integral part of instruction, its planning must go hand in hand with instructional planning, and the results must be used continuously to guide the planning and development of the curriculum."[1]

[1] Tyler, R. W., "The Functions of Measurement in Improving Instruction," Chapter 2 in E. F. Lindquist (editor) *Educational Measurement*, Washington, D. C.: American Council on Education, 1951.

"Teachers Don't Know How to Test"

JOHN A. GREEN
College of Education, University of Idaho

HOW MANY TIMES have you written in a qualifying statement on an objective test? How often have you failed to understand the instructor's meaning on a test? How often have you taken true-false tests with fewer than 20 items? How many times have you had the instructor casually wander into the room, announce a test, and proceed to write a discussion question on the board, a question which would require many hours of research and writing for a reasonably complete answer?

The number of questions which you answer affirmatively will give you some indication of the variety of measurement abuses to which you were subjected as a student. And these are but a few of the testing abuses which some high school teachers are imposing on their students. Unfortunately in the process of educational evaluation the test was seized upon by teachers as a means for justifying some kind of letter or percentage grade which was periodically reported to the parent instead of being used primarily as a diagnostic and measurement tool to facilitate the teaching of individual pupils and to measure individual attainment of the classroom objectives. Perhaps for this reason the test has often degenerated into a device for fooling pupils, a puzzle instead of a measurement tool. And the teacher may have emphasized in her tests a number of minor facts which were not mentioned in class or mentioned only in a cursory way. These teachers, the measurement and evaluation illiterates, have failed to note the fundamental principle that teaching objectives and measurement objectives are inseparably tied together and that measurement exists for the purpose of enlightening the teacher and pupil concerning that pupil's attainment of the classroom objectives. Such measurement carries with it the obvious, through frequently slighted, mandate that teachers state, preferably in writing, their teaching objectives. This statement of objectives in fundamental to a good teaching job since the teacher and pupils need to have definitely in mind the goals which they hope to attain, otherwise we have the case of the teacher who doesn't know where she is headed trying to measure pupil progress in many undetermined directions.

TESTING PRINCIPLES

There are several testing principles which can help the secondary teacher avoid some of the pitfalls cited above. The first of these principles has already been hinted at, that *the measurement objectives should be the same as the teaching objectives.*

Second, in most subject areas *it is preferable to have several types of testing instruments rather than one.*

Allied to this is the third principle, which requires that the *teaching objectives dictate the means of measurement.* For example, the informal objective test of the true-false or multiple-choice type is particularly useful if the objective is to measure the student's recognition or recall of a broad range of specific facts. On the other hand, the essay type examination is most useful if the objective is to measure the depth and scope of knowledge in a specialized, advanced field.

A fourth principle is to have the test *long enough to be valid and reliable but short enough to be usable.* A true-false test with less than 30-40 items is unreliable because there is too much guessing factor involved in the score, and the student on a subsequent retest might get an entirely different score. If the test is made too long, however, students cannot complete it within the time limits of the regular class period. Ordinarily an average high school student can complete three true-false items per minute. If an essay test is being used, increasing the number of questions improves the test by increasing the measurement sample; but, again, too many questions to be discussed in the class period may invalidate the examination by putting emphasis on writing speed rather than classroom objectives.

Fifth, the test questions or *items should be concise and understandable* so that they do not get in the way of measurement since they are merely means for conveying the teacher's meaning to the student. Obviously if the student misunderstands the item, he will be unable to give an acceptable answer even though he might know the answer if the meaning of the item were clear. It should be noted that objective test items which are stated negatively are often misunderstood by students but they have some value for screening the good from the poor readers provided that the teacher has the objective of measuring reading ability.

Sixth, the *difficulty of a test should be determined by its function.* There are three general types of tests: teaching, mastery, and measurement. Each of these test types has different functions and

a different level of difficulty. Teaching tests may be graded by students in class who have as their primary aim identification and correction of student errors, but they should not be used for assignment of marks. The mastery test is used to measure essential fundamental skills which each student should master, and almost all students might be expected to get all test items correct. Measurement tests should have about a 50 percent level of difficulty, and the items should be arranged from zero level of difficulty (answered correctly by all students) to one hundred percent level of difficulty (answered incorrectly by all students).

Seventh, *the test should include specific instructions and should be typed and mimeographed or dittoed* so that each student will have a copy.

TEST CONSTRUCTION

Four steps of test construction are recommended by Ross.[1] These steps are:

1. Planning the test
2. Preparing the test
3. Trying out the test
4. Evaluating the test.

The essential steps in planning the test include: (1) listing the objectives which are to be measured; (2) weighing the objectives to determine the importance of each in the finished test; and (3) deciding the type of test or tests which will give an adequate measure of all objectives. If these planning steps have been carefully executed, the task of preparing a valid test is infinitely easier than if it is approached in a haphazard manner.

Since preparation of the objective test is more difficult than preparation of essay tests, and since the true-false is the most frequently used teacher-made objective test, it has been selected to discuss here. Preparation of true-false tests becomes less of a task if the teacher prepares several items each day as she plans her classroom activities. These items should be copied on three-by-five inch file cards so that they can be filed by topic or objective for future reference. In preparing the items these rules should be observed:

1. Statements should be entirely true or entirely false.
2. False statement should be fundamentally false instead of being made false by a triviality.

[1] C. C. Ross and Julian Stanley, *Measurement in Today's Schools*. New York: Prentice-Hall, 1954, p. 140.

3. Verbatim quoting from the text should be avoided since it promotes rote learning.
4. A consistent pattern of answers should be avoided.
5. Quantative terms such as *ten* are preferable to qualitative term such as *few*.
6. Negative statements should be avoided.
7. Statements should be no longer than is necessary to clarify the teacher's meaning.

Sufficient time should be allowed in the preparation of the test to permit it to be typed, proof read, and duplicated. After the test has been tried out for the first time, a simple item analysis can be made by counting the number of correct responses on each item by students in the top third and the bottom third of the class. This information can be entered on the file cards on which the item was originally written, and it gives the teacher a basis for refining the item for later use. Over a period of years a teacher who prepares her tests carefully and keeps the items and item analysis data can build up an extensive file of good test items from which several non-duplicative tests can be prepared to measure a given set of teaching objectives. Careful, planned measurement and evaluation can pay dividends for every student in the classroom by providing the teacher with individual blueprints on the attainment of classroom objectives.

Measure—Evaluate—Then, Adapt

VIVIEN INGRAM
Consultant in Guidance and Testing
SUSAN KILMER
Trainee Assistant in Communications, Flint (Michigan) Public Schools

TEST RESULTS are useful only if used intelligently, and evaluation is of value only through intelligent application. The Flint community schools have an intensive rather than an extensive educational testing program. It is their aim to use each pupil's test results as completely as possible. They believe the program has value only if its results aid teachers and administrators in two areas. First, test results should aid in understanding the child so the teacher can help him to adjust to school environment; and secondly, results should indicate where changes in curriculum would help the child and the whole school program.

To understand the child, teachers and school officials must become more intimately and reliably acquainted with the educational accomplishments and capabilities of each pupil. Through appreciation of a pupil's talents and limitations, instruction and guidance can be better adapted to his individual needs, interests, and abilities.

In the Flint secondary schools—grades seven through twelve, five tests specifically measure the child's achievement and ability and two additional tests provide information for individual counseling:

In the seventh grade, the California Mental Maturity test is used to measure mental growth with separate evaluation for the verbal and the non-verbal pupil.

The Science Research Associates Achievement test (SRA) given in the eighth grade is a practical tool in helping to adapt instruction and guidance to individual needs since it provides measures of basic skills. This battery measures as directly as possible the objectives of the entire school program. It measures skill, not content, in the tool subjects.

At the tenth grade level, the California Test of Mental Maturity helps to measure mental growth for the academic and non-academic pupils. This measurement gives counselors and teachers aid in determining whether a pupil is working to the limit of his capacity.

The Iowa Test of Educational Development, given in the eleventh grade, measures fundamental skills which have been developed both inside and outside the classroom. It attempts to measure the influence of instruction.

The Gates Reading Survey is used as needed in secondary schools to determine placement in special reading classes.

Tenth grade pupils receive the Kuder Interest Inventory which gives a systematic record of preference in a variety of activities and is helpful in the selection of occupations to investigate and consider for the pupil's life work.

Seventh and tenth graders fill out a background information sheet to bring together all essential information about the pupil so that all test results can be interpreted in terms of his needs.

Test results are used in counseling the individual and as a means of evaluating an adapting curriculum. They help the counselor to understand the child and the child to understand himself so he can learn to adjust to his needs.

By examining many test results over the years, many changes have been made in curriculum. These changes have been at the secondary level in the areas of English, mathematics, social studies, business education, and personality and character development. In addition to course changes, many new fields have opened up; for example, the education of the talented child with emphasis on mathematics and science, the experimental low achiever program for pupils of normal intelligence who are below grade level, inovations in the program for the mentally handicapped, and a new concept in administrative organization of the junior high school or the "house" plan.

Two classes in English have been added at the twelfth grade level. A combination social studies and English class with a two-hour block of time has been established for superior students. Also, for the academically talented, a one-semested class in composition and grammar accompanies the usual one-semester class of English literature.

In tenth grade, talented pupils are offered a course entitled *Humanities* which gives English credit with science or social studies credit depending on the pupil's interest and the nature of his project work. The pupil is encouraged to plan and execute his own

work. For the talented child at junior high level, the emphasis is on science and mathematics.

In the area of mathematics, two tests are used. The Schorling-Clark-Potter 100 Problem Test of Arithmetic Fundamentals is used in grades seven, eight, and nine each year. If test results indicate that a pupil has not reached a desirable level of achievement by the time he has completed the ninth grade, he is placed in a Basic Mathematics course in the tenth grade.

The second test used to evaluate mathematical ability is the Orleans Algebra Prognosis test administered in the eighth grade. This test, in addition to other information gained from the pupil's mathematics records, provides a guide for counselors to determine whether a pupil should attempt algebra in the ninth grade or should continue with fundamental arithmetic processes for another year.

By the twelfth grade, many pupils have a personal need for practical mathematics. These young people are usually not prospective college students and have not elected algebra or geometry or other advanced mathematics courses. This special mathematics class gives information about tax computation, borrowing, lending, and similar everyday applications of mathematics.

Because of the great number of pupils whose test results show a need for help in reading (as evidenced by Gates Reading Survey) and mathematics (Orleans Algebra Prognosis and Schorling-Clark-Potter test), an experimental low achiever program which teaches remedial mathematics is now in operation. Pupils who need extra help meet every day with a special teacher in place of their regular mathematics or reading class. When the child reaches a level of achievement where he can successfully meet the requirements of his grade, he resumes his mathematics or reading with his regular class.

A summer school program, from grade three through grade eight, for remedial reading and remedial arithmetic has replaced the conventional summer plan of repeating all subjects failed. Small classes meet for two hours daily for six weeks. This has proved to be a highly successful way of helping pupils with weaknesses in the basic skills.

For those who are mentally retarded, Flint schools offer classes for the educable pupil who is socially competent. If physically and mentally able, twelfth grade pupils are placed on a job in the community. These young people take as many courses in the regular curriculum as they are able.

A recent followup study of Flint high school graduates revealed that pupils felt they received the least value from social studies classes. The followup study seemed to indicate that this subject matter was too remote from their needs to be of genuine value. This evaluation was an influence in the establishment of an entirely new class primarily for pupils not enrolled in college preparatory work. A course entitled *American Horizons,* designed for the student whose range of experience is limited, is offered at the tenth grade level in addition to the ancient and world history course which is offered for the college preparatory students.

At the twelfth grade level, an *American History and Civics* course is now taught for non-college pupils, in addition to the college preparatory *United States History and Government.*

In order to stimulate interest in and increase knowledge of our American ideas and ideals, the Flint Board of Education-owned-and-operated FM radio station, WFBE, presents a series of programs for use by social studies classes. The taped radio program series "When Men Are Free," written and produced by Station WHA of the University of Wisconsin, is used specifically to make the ideals of American thought and government more vivid.

Another example of a radio program's helping to fill a need is a series which helps the teacher stimulate good creative writing.

A guidance series, written and dramatized by the school radio station staff, is effective in dealing with abstract qualities of character development, such as honesty and cooperation.

In the area of business education, several new programs have met not only the needs of the individual, (as in the Cooperative Occupational Training and Placement Services) but also the program of the industrial curriculum has been aided and encouraged by industry to help meet the needs of the community.

The Cooperative Occupational Training Service is available to students in distributive education, trades and industry training, and business education. Pupils of 16 years of age or over, mentally and physically able to benefit from this program, are given the opportunity of work experience in the community along with a school-related subject to correlate the actual experience with classroom instruction related to the pupil's job.

The industrial curriculum was a cooperative attempt by leaders of Flint industry and staff members of the Flint Public Schools to raise the standards of the existing industrial arts curriculum.

Units in report writing were added to English classes and a

course was developed in the *Economics of Our Community* to meet the needs of men in industry. Pupils are urged to take advanced mathematics and additional science classes to equip them for our modern technological society.

A *Day Trade Auto Mechanics course* was initiated at one high school to prepare pupils for apprenticeships in auto mechanics and collision work. This program provides three hours a day on the job, supplementing the school correlated program.

Several special classes have been developed to meet special needs. A typing class of one semester gives those pupils who haven't time to elect the usual one-year course an opportunity to learn to type. Many college preparatory pupils find this course especially practical. Also, pupils training in distributive education to perform services, such as a waitress, may find a short typing course a practical help.

For the shy and retiring pupils and for those with speech defects of a type which would respond successfully to group therapy, a public speaking class is offered. The instructor attempts to build poise and to develop correct standards of speech.

The Flint schools feel responsible for the education of the whole child, and tests, as well as other evaluation techniques, have revealed weaknesses in areas of pupil self-knowledge, pupil relationships with his family and his community. A new course was established at the seventh grade level entitled *Effective Living at Home, School, and in the Community.* This course attempts to help the child understand himself and those about him.

A similar course, elective at the twelfth grade level, entitled *Family Living,* has much the same aim as the seventh grade course but is designed for advanced students.

Flint school leaders have felt for many years that the conventional junior high school has been too often a "little senior high school" and as such, has not met the emotional, social, and academic needs of the junior high school pupil. School administrators realize this age group needed more guidance. The "house" plan now in an experimental stage in two junior high schools helps solve the problems which arise from a lack of personal contact between teacher and pupil.

This year two large schools of 1800 or more pupils were theoretically divided into three smaller school sytems or "houses" within the larger system. A team of eight teachers in each "house" is responsible for teaching the academic subjects and for the personal-social-emotional adjustment of the students in their "house."

One teacher is responsible for one group of pupils in the morning and another in the afternoon, teaching two subjects, such as English and social studies, and for the counseling of both groups. The core team teachers work together exchanging information about pupils in their classes to build a more effective guidance program. In order to increase the possible effectiveness of the teacher as a guidance person, he is relieved of his clerical and non-professional duties by a non-professional teacher aide, who serves all eight teachers on a core team.

These examples have been some of the obvious changes in curriculum. They are not the only changes. As every creative teacher uses his test results to improve his instruction, adaptations are often made.

Constant evaluation, testing, and retesting are necessary to adjust the schools to the needs of the pupils. A static school curriculum does not fit in a constantly changing world. Material must be kept vital so each child is assured of an education to the limit of his ability which prepares him for the era in which he is living.

Can We Measure What We Teach?

JOHN C. FLANAGAN
American Institute for Research and University of Pittsburgh

OUR ABILITY to measure what we teach has improved substantially during the past twenty-five years. During this period the distinction between what the student has learned in terms of specific information and his understanding and ability to use what he has learned in new situations has been substantially clarified. This trend is clearly evident in the Cooperative Achievement Tests, the Iowa Tests of Educational Development, and the Sequential Tests of Educational Progress.

This is a very important distinction and teachers are advised to check the many interesting and novel types of items prepared by state and national committees for these tests. It takes something more than a test which measures the student's knowledge of names, dates, and events, to measure his understandings of contemporary social institutions and practices. For example, the sample item from the Understanding of Basic Social Concepts of the Iowa Tests of Educational Development measures the student's understanding of a trend rather than an isolated event.

How does the importance of machines today compare with their importance in the past?

1. We make more use of machines than people have at any time in the past.
2. People have made about the same amount of use of machines ever since the factory system was developed in England in the latter part of the 1700's.
3. Machines reached their present stage of importance about the time America was discovered.
4. Civilized people have always used machines about as much as we do today.

Similarly, the Sequential Tests of Educational Progress Mathematics Tests are constructed to assess competence and skills essential in the recognition and application of six important concepts. A sample of this type of test item is as follows:

Situation: Mr. Jones has a dairy farm on which he also grows corn.

Mr. Jones said that 3/4 of his cows were Jerseys, but only 2/3 of his neighbor's cows are Jerseys. If the neighbors herd is larger than Mr. Jones', which one of the following statements is true?

A. They have the same number of Jerseys.
B. Jones has more Jerseys.
C. The neighbor has more.
D. It cannot be determined who has more Jerseys.

It is clear that this item also measures something more than ability to perform mathematical manipulations.

Substantial progress has been made in developing objective items to measure Writing Skills. An example of some of the types of questions included in the Writing Tests of the Sequential Tests of Educational Progress are given below:

My Favorite Magazine

1 Many young people of today are taking a great interest in the magazine *Suburbia*. 2 It is of a fairly large size with a considerable number of pages. 3 The publishers, Allan and Watts, are well know and reputable; thus providing young homemakers with ideas and practical plans for their present and future homes. 4 Because it also contains articles of family and community relations and hints on home improvement, it is most likely preferable reading to people who are seeking guidance on such mattters. 5 The previously mentioned content explains why the advertisements would logically be about products and furnishings for the home. 6 The articles were well written, and all the features of *Suburbia* help to form an interesting and informative magazine. 7 I found one particularly interesting article, it was entitled "Families Are Using Spare Time to Broaden Their Horizons." 8 This article points out that leisure time should be spent in any activity other than usual work, instead of remaining in complete inactivity.

8. In Sentence 2, how could the size of the magazine be indicated most effectively?

E. By comparing it with one or two well-known magazines.
F. By giving length, width, thickness, weight, and number of pages.
G. By drawing a scale model.
H. By telling how many articles each issue contained.

. . . . .

11. Sentence 5 is awkward. Which of the following revisions is best?

A. Advertisements are about products and furnishings for the home because this is logical in a magazine of such content.
B. The advertisements would logically be about products and furnishings for the home, like the articles.

C. In keeping with the content of the articles, the advertisements are about products and furnishings for the home.

D. Because the content of the articles, as previously explained, are about homemaking, so the advertisements properly are also.

---

14. Which of these possible additions to the report would be most helpful

E. A listing of the advertisers making use of several issues of *Suburbia*.

F. A statement of the rates paid by the publishers for articles purchased.

G. A listing and brief description of the "departments" appearing in every issue, as distinct from the featured articles.

H. A statement, accompanied by authoritative references, of the credit rating of the company which publishes *Suburbia*.

The above items include many important aspects of the active process of expressing ideas in writing in the form of objective test items. It is, of course, not suggested that such items replace the student's preparation of essays. It is suggested that such items provide a standard methods of pin-pointing writing errors which should be of great value to the teacher.

Similarly, we have learned to measure the student's understanding, interpreting, and evaluating of typical materials in the field of natural science or social studies. We have also developed some very effective procedures for measuring appreciation and comprehension of various types of literary materials.

What are some of the areas in which our measures have not kept pace with our teaching? One area in which there has been considerable lag is that of measuring our ability to develop creativeness and ingenuity in our students, particularly in the field of science. There has been considerable activity in this field recently and it seems likely that a satisfactory method of measuring the teacher's success in this field should be available shortly. One of the types of items measuring this objective is illustrated by the following sample of the Ingenuity Test of the Aptitude Classification Test Battery:

A very rare servere wind storm destroyed the transmission tower of a television station. The station was located in a town in a flat prairie with no tall buildings. Its 300 foot tower enabled it to serve a large farming community and the management wanted to restore service while a new tower was being erected. The problem was temporarily solved by using a

A. p---n
B. k---g
C. c---e
D. b---n
E. t---r

These items measure the student's ability to reproduce an ingenious solution to an important problem in a practical or scientific field.

Perhaps the areas in which our present measurement procedures lagged the most include some in which we profess most interest and concern. These include such objectives as character development, citizenship, and social adjustment. Unfortunately, our measures of the attitudes, habits and social skills have not attained the same degree of precision that has been reached in most other fields of educational objectives. A considerable amount of research and developmental work is being devoted to these measurement problems and it is hoped that satisfactory measures will be available in the not too distant future.

Of the various approaches being tried at the present time the observational checklists, self-concept report forms, and situational performance tests appear to have substantial promise. The values of the first and last of these procedures to derive principally from their objective and systematic tabulation of actual behavior in a real or simulated situation. The unique value of the second type is in its direct apporach to basic motivational factors.

In summary, it appears that we have made substantial progress in measuring the intellectual aspects of what we are teaching. Much less progress has been made in measuring the non-intellectual objectives of education. It appears that the next decade should witness substantial progress in providing standardized measures of satisfactory reliability and validity in these areas.

REFERENCES

Cooperative Test Division, Educational Testing Service: *Sequential Tests of Educational Progress.* Princeton, New Jersey: 1957.

Flanagan, John C. *Aptitude Classification Tests.* Chicago: Science Research Associates. 1958

State University of Iowa. *The Iowa Tests of Educational Development.* Chicago: Science Research Associates. 1957.

Are Teachers Qualified to Develop, Administer and Interpret Tests*

DUGALD S. ARBUCKLE
Boston University

IN THE MANY arguments about modern teachers—what they have and what they do not have—one of the most common comments is something to the effect that the modern teachers knows too much about how to teach and not enough about what to teach, while the more ancient variety of teacher knows too much about what to teach and not enough about how to teach. Thinking in terms of a *good* teacher, in a modern American school, it is difficult to visualize him performing some of his most crucial duties if he knows little or nothing in the field of testing and measurment, which might be considered to be an aspect of the "how" of teaching rather than the "what" of teaching.

THE NEED FOR TEACHER UNDERSTANDING OF TESTS

The need for some teacher understanding of tests and other means of measurement would seem to be quite obvious, and yet it is true that not only do thousands of present-day teachers have little or no understanding of tests, but many others are being graduated as teachers without the slightest understanding of testing and measurement. Secondary school teachers, particularly, are sometimes quite ignorant about the people whom they teach, although they may be relatively well versed in what they teach. Why, then, should the high school teacher of mathematics have some understanding of tests?

He cannot do a good job as an academic teacher without being at least able to talk the language of testing. What, for example, does he use as a criterion of academic achievement for his mathematics students? Does he know anything about the various achievement measures that are available from numerous reputable, and possibly some less reputable, publishing houses, or does he depend on a guidance department, which does not exist in many schools? It is the responsibility of the teacher to take action so that he has some understanding of the camparison of the achievement of his students as compared on a local, a state and a national basis. In

*The role of the teacher in the use of test data is described in much more detail in: Arbuckle, Dugald S. *Guidance Counseling in the Classroom;* Boston, Allyn & Bacon, 1957.

a small school system, the teacher might well have to administer, score, and interpret such tests himself. In a large system, the teacher should know enough about achievement testing so that he knows what he wants, and he can make it clear that if the guidance department does not help in obtaining this information, it is not performing one of its vital functions.

Nor can the teacher do much of a job in helping the students to attain some understanding of the intricacies of mathematics if he does not have some understanding of their intellectual assets and liabilities. Even if we speak in terms of the much maligned IQ, the teacher should surely know the intellectual capacity of his students, he should know that in his class there is a difference of 47 IQ points between the highest and the lowest score, and he should know just what this means in terms of possible performance. He should also know the different types of tests, so that when he sees an IQ score which is the result of three indivdual Stanford Binet's, administered by a certified psychometrist, he may be aware of the fact that this may not mean the same thing as an IQ score which is the result of a group Otis, administered by a teacher to a group of 200 students. Even though the teacher may work in a somewhat selective school, so that the students in his mathematics class have already gone through a screening process, he would be naive indeed if he assumed that his students had a somewhat similar degree of academic capacity and background. He would probably be equally naive if he assumed that since a student was in his class in mathematics, he must naturally have the intellectual capacity to do a decent job in that class.

The pupil's achievement is not, of course, geared solely to the student's intellectual potential. There are many bright students who are having much difficulty in academic achievement, and here again the teacher is going to be handicapped if he does not have some understanding of the various factors, other than intellectual capacity, which may result in underachievement, and if he is not aware of some of the measures which may be used to at least attempt to get at this problem. Even the traditional teacher, who may believe in mathematics for the greater glory of mathematics, would probably agree that he can do a better job of teaching mathematics if he knows something about those items which go to make up the total personality of the students whom he is teaching.

The teacher cannot function as a real *teacher,* rather than an instructor, if he does not know something about the personality make-up of his students. If the teacher is to be really concerned

with the total development of the child with the social, the emotional, the religious, the avocational side of his personality as well as his intellect, if he is really concerned with the child as a responsible member of a free and democratic society, if he is really thinking in terms of the possible contribution of the student to his society and to his fellows, then he must surely have some understanding of this total student, entirely apart from any relationship that might have to academic proficiency. The secondary school teacher in our modern American society surely cannot see himself and his students as living in an isolated academic oasis. And, if he does see himself and his student as being of necessity concerned with many factors other than the academic life of the school, then he should be aware of those instruments which may give to the teacher a better picture of the total personality of the child.

Another obvious reason for the necessity of teacher understanding of testing and measurement is the fact that practically every teacher is involved in the construction of tests, quizzes and examinations which supposedly indicate to him the progress and understanding of his students. Since the teacher must develop his own tests, it would not seem unreasonable to say that he should have some understanding of test construction.

THE DEVELOPMENT OF TESTS

The teacher can hardly become involved in the development of tests without some rudimentary knowledge about testing and statistics. How many teachers, for example, can give some statistical evidence that the tests they use are both reliable and valid. If an English teacher uses a certain test, the student has the right to feel that the scores that are given to him will be the same if the test is corrected on November 3 as if it is corrected on November 27; the same if the test is corrected by either Mr. Jones, Mr. Brown, or Miss Pell. Consistency between different teachers may be too much to hope for, but the teacher can hardly be satisfied with a situation where there may be a significant difference in the score on a paper if it were corrected by Mr. Brown instead of Mr. Jones, and an even greater difference if corrected by Miss Pell. One teacher may say that he cannot do much about the consistency of correction among different teachers, but he should be able to show some consistency in his own correction of a certain paper. In some subjects, such as English and Social Studies, there is much more likelihood of a low reliability than in such subjects as mathematics or chemistry, where the personality of the teacher is less likely to affect the grad-

ing of papers. Thus the teacher who develops his own achievement tests must be able to indicate the degree to which these tests are reliable. A student should be able to assume that he will get the same score on his test paper, regardless of when, or by whom it is corrected.

An achievement test and most tests developed by teachers are measures of achievement and should also be valid, in that it measures what it is supposed to measure. An algebra test can hardly be described as valid if the teacher states that "John can hardly expect me to give him a good grade with the sort of attitude that he has . . ." An algebra test would, one might assume, measure the student's knowledge of algebra, not the attitude of the child, although his attitude may, of course, affect his understanding of the subject under discussion. Thus, the teacher must use measures to determine if his test measures exactly what it is supposed to measure. Is it a certain area of knowledge discussed in several books, in lectures, in class discussions? Does it cover work taken up in the past month or the past six months? Is it a measure of the child's knowledge to think and reason, or rather his degree of retention of data? Many students complain, with justification, about the lack of validity of a test. They do not, of course, use this word, but this is often what they mean when they utter various gripes about tests, such as, "He never talked about that," or, "He told us to study the book and there wasn't a single question from the book," or, "How can he expect us to remember all that junk," or, "How can he expect us to understand a stupid question like that," and so on.

Most of the tests which are used by teachers cannot be used for comparative measures, and this is a major reason why every teacher should use certain standardized measures as well as his own local measures which usually only compare the child with previous or similar classes in the school. If the test is not reliable and valid, of course, it does not even do this.

Many professional teachers may become interested in developing tests which will measure certain traits or characteristics or attitudes for which they feel no adequite measure exists, and they will become involved in a research project which will result in the construction of some new measurement instrument or device. This, however, calls for a good deal of research "know how," and no teacher should assume that he can casually whip up some testing instrument which will be likely to be worth any more than the paper on which it is written. The greatest initial contribution that teach-

ers can make in the development of tests is to see that the tests they use to measure the achievement of their students have a high degree of validity and reliability.

THE ADMINISTRATION AND INTERPRETATION OF TESTS

A vast array of tests and various types of measurement instruments are available for teachers. Probably the majority of them are group measures which can be administered quite effectively by a teacher who is intelligent enough to read and to follow the directtions. Any of the reputable test companies are glad to put teachers on their mailing lists, so that any teacher can have a fairly up-to-date indication of just what is being produced in the field of testing. On the other hand, there are certain tests that cannot be administered unless the teacher has some professional preparation for the task. The Revised Stanford-Binet Scales, for example, is an individual measure of intelligence, certainly one of the most respected of testing devices, but one must take certain course work from an accredited institution which offers programs in this area before being certified to administer such a test. The Wechsler-Bellevue Intelligence Scale would be another measure of intelligence which needs professional preparation so that one may be capable of administering the test. Most projective tests such as the Rorschach, should also be administered only by those who have a professional background in the field, and most test companies have certain instruments that they send only to those who are professionally competent. On the whole, however, it would be safe to say that the intelligent and interested teacher is quite capable of administering the vast majority of standardized tests that are published today.

One must, however, enter a word of caution at this point. Every teacher who is planning to administer a test should be certain as to just what his objectives may be in the administration of the test, and quite frequently teachers will apparently feel that they are going to get from the test information that which would cause the author and the publisher to throw up their hands in horror and say, "That's not what you'll get from this test." The teacher, for example, may want a test which will tell who, among her students, will be successful scientists. There is no test published which will give her an answer to this question, although there are various kinds of tests which may give her information which will help her in finding a more accurate answer to her question. Teachers should also be aware that no test is completely valid and reliable. There are

Boston University
School of Education
Library

always students for whom test data is quite invalid. There will always be very intelligent children who, according to an IQ test are sub-normal; there will always be children who appear, according to test data, to have little, but actually they have much. On the whole, achievement and intelligent measures are much more valid measures of what the child *has* rather than what he *does not* have. If a person can run 100 yard in 12 seconds we *know* he can run the 100 yards in 12 seconds; we do not know, however, whether or not he can run it in 11 seconds. Tests, then, particularly intelligence and achievement tests, tend to indicate what we have rather what we do not have. Measures of attitudes and personality traits, on the other hand, are generally more questionable measures of both what the student has or what he does not have since these items are not as clearly defined and are not as easily measurable. A Stanford-Binet measure which says that the mental age of a child is 10 years 6 months is more likely to be a valid picture than is the result of a personality measure which indicates that the child has poor personal relations with others.

In most cases, an individual teacher who is capable of administering a test is also capable of interpreting the results, as long as they operate within the frame of reference that is usually described in some detail by the author of the test. Thus some teachers who use a Kuder Interest Inventory may tell a student that the test data indicates that he should be successful as a teacher. The Kuder gives no such indication, but it does indicate how the interests of the student compare with a sampling of successful teachers; an intelligence measure does not indicate whether or not a high school student will succeed in college, but it does give some indication of the degree to which he has the intellectual capacity to do the academic work which may be expected of him in college; a finger dexterity test does not differentiate between a student who will do well in typing and one who will do poorly, but it does indicate that one student has more of a certain skill or capacity that is usually necessary if one is to become an efficient typist. Thus if teachers read test data carefully, and if they are cautious in not going beyond what the test was constructed to do, they should be able to interpret most of the tests that are available today. This does not, however, include projective devices, some personality inventories, and certain individual tests such as the Stanford-Binet. Certainly, in modern day education, however, the high school teacher who is to perform his duties in an efficient manner must have some understanding of tests and other instruments of measurement.

Sociometric Measurements in the School

WALTER D. SMITH
Department of Psychology and Child Development Institute
Florida State University

THE TEACHER who has embraced the "whole child" concept and follows-up his philosophical convictions with positive attempts to bring reality to these convictions may include in his teaching techniques the sociometric test with its matrix sheets, sociograms, and perhaps some statistical analyses, all calculated to improve the lot of children. The practical question, "What does the sociometric test measure?", may be less understood by this teacher than the administration, scoring, and statistical analyses used. The meaning or significance of the sociometric analysis of the classroom requires a broader understanding.

The sociometric test is designed to measure children's social status and the social structure of a group. It reveals only a single facet of the classroom situation, a cross-sectional picture of the social scene which may shift and change in response to environmental forces before the next measurement takes place. What is measured and what it means can be explained only by determining the relationship between sociometric test scores and certain external variables. Evidence of validity is thereby provided, thus enhancing the interpretation of the sociometric data. For purposes of achieving the maximum illustrative evidence, children falling on the upper and lower ends of the social status continuum as derived from sociometric testing may be compared on the basis of these external variables.

Children on the Extremes of the Social Status Continuum

The social "stars," the children of highest social status in the classroom, have been characterized as: more intelligent, more mature, more active, superior in athletic skills and general health, more attractive physically, more sensitive to their social environment, and more able to react conventionally and to control their emotions in social situations. These findings seem to appear, even when cultural boundaries are crossed. A recent study by Thorpe (1) showed that the findings summarizing the relationship between social status and personality in American school children appear in almost identical form in similar studies of English school children.

The social "isolates," on the other hand, displayed behavioral characteristics observably different from those describing children of higher social status. They have been variously described as: recessive, listless, of low vitality, disinterested, of poorer health, shy, bored, socially ineffective, rebellious, noisy, of lower intellectual ability, less competent academically, less attractive physically, boastful, and arrogant. Personal problems are reported to occur more frequently in this group which is also characterized by many writers as being of lower socio-economic status. With older children and adults the descriptions may be differently phrased. The least chosen individual may be persuasive, confident, and dynamic but at the same time he may not be desirable to work or play with because of such characteristics as being overly competitive, or not being flexible and logical in his behavior. These descriptions and kinds of behavior have been variously accounted for as efforts to overcome insecurity and social isolation, as due to innate temperamental differences, or to other predisposing factors.

From the reseach evidence it may be concluded that high social status in children's groups shows a positive relationship to success in various areas of performance and to personality adjustment and maturity, and that the reverse picture is presented by children of low social status. However, in every situation it must be noted that no single trait has been found to be all-important. Compensatory positive traits may completely overshadow a variety of less desirable traits and thereby result in the child's achieving high social status. Berkowitz (2) showed that prior social behavior, as social ascendancy, may be far less important than situational factors which appear to elicit from or draw-out of the individual responses necessary for his optimum social survival. It becomes apparent that the measurements derived through sociometric testing require both a personality and a situational analysis for fullest understanding. The teacher's first step toward sociometric testing leads almost inevitably to other child study techniques.

Reliability and Validity

The reliability of the sociometric test appears to have been sufficiently established to anticipate a persistent constancy in the individual's social status from one time to another. A recent analysis (3) of 53 reliability studies showed the cumulative evidence from three decades of research to firmly support the reliability of the sociometric test, with reliability coefficients increasing with in-

creasing age of children's groups. A similar review (4) of validity studies by the same authors showed that social status is predictive of such performance criteria as productivity, leadership, training ability, and combat effectiveness and that the evidence is such as to encourage the use of and research with sociometric methods as predictors of other performance criteria.

Social Status and Group Behavior

Individuals who work and live together from differentiated subgroups, cliques, and reciprocal attraction or rejection relationships, all of which, in turn, produce marked effects on work efforts, group cooperation, individual effort and initiative, and group morale. An army barracks, the football team, the college sorority, and the high school clique or friendship group may be examples. Research in both education and industry suggests that these social relationships are far more important than are such incentives as increased pay, living conditions, and other material rewards from teachers or leaders. An educational system designed for teaching and learning in a group setting provides such a complicated variety of stimulus conditions to be sorted out, labeled, and measured that teachers and social psychologist must be ever restless, curious, probing individuals; their unit of study is always, in the final analysis, this complicated individual, immersed in the infinitely more complicated group.

Each school boy or girl differs from each of his peers in some degree in all his growth characteristics, and, just as importantly, he differs in his precise social status in the group, in the bonds of acceptance between himself and every other group member along which ideas seem to flow and influence is exerted. One might question where the heart of education is located, in the group or in the teacher? Even the teacher must find his place in the group.

Sociometric choice represents a conscious decision on the part of one individual to structure the social field and to accept and to collaborate with another person in future action. An implied, emotional bridge then exists across which the least inhibited sort of expression and communication can flow. Spontaneity theory of sociometry indicates that the most imaginative, resourceful, problem-solving efforts will emerge in this atmosphere. Is this what we want in education? Newer concepts of democratic practices in the classroom bring these issues into the open. The crises of daily living in the United States with the ever present fantasy solution of going back to the older, tried-and-true techniques force all thinking

people into problem solving behavior which frequently arrives at the answer: "We can accept this point of view with the particular kinds of developmental problems it promises because we see the matter in its long range perspective; today's children must become tomorrow's resourceful adults and must grow toward that goal by learning to make choices early in life.

Sociometric Measurement and Spontaneity Theory

Moreno (5) stated that behind the whole theory of sociometry was his concept of spontaneity. Moreno's view emphasized the freeing of the person to meet each new situation with adequacy through the helping role of other persons who could assist in freeing the individual. A degree of spontaneity is always necessary before the individual can react at all to other persons. It might be argued that inventions and all creativity are the products of spontaneity. A next question to be answered by educators, then, is, "When children are permitted to structure their social environment in accordance with preferences expressed in sociometric tests, will the behavior which then occurs be more spontaneous, freer, more creative, and will it bring about more positive interpersonal relations?"

Wieland's research (6) carried out first in a pilot study with sixth graders and then repeated with kindergarteners offered some support for Moreno's views. After initial sociometric testing in the sixth grade classroom, child A was paired with his mutual choice, child B, in an experimental play situation where moulding clay and pipe cleaners were available for construction of various design and structures (these two children had chosen each other). On another occasion child A and child C were paired in the same play construction situation (neither choice nor rejection were expressed between these two children). Observations were made of all behavior and sound recording were taken. Then, child A's behavior and his play product from the two experimental situations in which he served as subjects were compared. When he worked with a mutual choice, child B, spontaneity would be expected to be high; when he worked with a child C for whom no choice or rejection was expressed, spontaneity would then be expected to be low.

The statistical analysis of Wieland's study, while not uniformly conclusive, generally showed child A's behavior to more frequently fit into the "dominative" category, as described by Anderson, (7) when he was paired with the unchosen child C. Poorer work products and frequent differences of opinion were characteristics of this pairing. Creativity in his art product and integrative behavior in his social

relationships were always more typical of the child A-child B pairing than of the child A-child C pairings. Observational reports of the two sessions showed child A to be task-occupied for a longer time, to be more often engaged in task-oriented discussions, to exchange ideas and help more freely, and to be less frequently disturbed in his work and thinking about his project when he was paired with his mutual sociometric choice.

Meaning of the Evidence

Our mental health goals and our schools' task-oriented goals favor the consideration of the child's help in shaping the social world of the classroom. The evidence is consistently favorable—from public schools, industry, penal institutions, scouts, university classrooms, the military, and a host of other sources. The sociometric test, when properly used, provides the teacher with a measure of the quality of social interaction within the classroom. Children's learning and adjustment are influenced by and perhaps even products of this social interaction. Neither adult, adolescent, nor child takes these social relationships lightly. In the long run it would appear that boys' and girls' social relationships represent youthful investments in one of the most stable values of our time, other people. For some two decades sociometric tests have served to enhance the helping role of teachers by providing a better glimpse into this social investment our children make and experienced teachers have often demonstrated the utility of these measurements in the everyday practices of the classroom.

BIBLIOGRAPHY

1. THORPE, J. G. An investigation into some correlates of sociometric status within school classes. *Sociometry*, 1955, 18, 49-61
2. BERKOWITZ, LEONARD. Personality and group position. *Sociometry*, 1956, 19, 210-222.
3. MOUTON, JANE SRYGLEY, BLAKE, ROBERT R., AND FRUCHTER, BENJAMIN. The reliability of sociometric meaures. *Sociometry*, 1955, 18, 7-46.
4. ———————. The validity of sociometric responses. *Sociometry*, 1955, 18, 181-206.
5. MORENO, J. L. AND MORENO, FLORENCE B. Spontaneity theory and child development. *Psychodrama Monographs*. 1944, 8, 42.
6. WIELAND, R. G. An examination of the spontaneity theory of sociometry. Unpublished Masters Thesis, Florida State University, 1953.
7. ANDERSON, HAROLD H. Domination and integration in the social behavior of young children in an experimental play situation. *Genet. Psychol. Monogr.*, 1937, 19, 341-408.

Recent Books in Measurement*

WILLIAM P. MATTHEWS

Measurement and Evaluation for the Secondary-School Teacher, by Georgia S. Adams and Theodore L. Torgerson. New York: Dryden Press, 1956. 658 pp.

The authors in this book have attempted, and, I might say, have done an excellent job of trying to get the teacher to recognize the principles involved in learning. Following this they have focused on the implications of an understanding of these principles. Then they have recognized the tremedous effect in the teaching of the mass of students who will be coming into the school and the difficult job of understanding the individual differences of these students. In the final anaysis it is both the student and teacher who will benefit from a better understanding of the students' abilities and performance. This book helps to bring out the fact that we need to prevent learning difficulties as well as to detect and correct them if we are to provide for the need of students. The book is divided into four parts: The Evaluative Process, The Study of Individuals, The Improvement of Instruction, and The Administrative, Supervisory, and Guidance Aspects. W.C.M.

Measurement and Evaluation in Psychology and Education, by Robert L. Thorndike and Elizabeth Hagen. New York: John Wiley & Sons, Inc. 1955, 575 pp.

While this is not the most recent book it is one that will be used for quite a while I am sure, because it is based on scholarly writing, a sound and seemingly thorough understanding of the subject by the authors, a necessary knowledge of the theories involved, and, most important, an interest in getting all of this across to potential users of the text. It is both a reference and a textbook that should have a wide range of appeal to teachers, counselors and psychologists because it is written, it seems, with their needs in mind. It is written, the authors say, as a general introductory text and as such it should be more popular than one that verbalizes in the theoretical stratosphere only. The book and its contents remind me of a good meal which one leaves thinking he would have like more but at the same time realizing that it was complete. W.C.M.

Psychological Testing, by Anne Anastasi. New York: The Macmillan Company. 1954. 682 pp.

Measurement and Evaluation in Education, by James M. Bradfield and H. Stewart Moredock. New York: The Macmillan Company. 1957. 509 pp.

Measurement in Education, by A. M. Jordan. New York: McGraw-Hill Book Company, Inc. 1954. 533 pp.

Measurement in Today's Schools, by C. C. Ross and Julian C. Stanley. 3rd Edition. New York: Prentice-Hall, Inc. 1954. 485 pp.

Educational Measurement, by Robert M. W. Travers. New York: The Macmillan Company. 1955. 420 pp.

*In reviewing books for this issue on measurement and evaluation we should like to emphasize that the books reviewed or listed here are not the only worthwhile books in this area. They are among the best.